Special Report
SR-WATSD-99-44

September, 1999

Third-Party Checking of 1999 Scaling and Linking for the Kentucky Core Content Test

R. Gene Hoffman
Arthur A. Thacker

Prepared for:

The Kentucky Department of Education
500 Mero Street
Frankfort, KY 40601

Contract Number BP 010492

Human Resources Research Organization
66 Canal Center Plaza, Suite 400 • Alexandria, VA 22314-1591

Third-Party Checking of 1999 Scaling and Linking for the Kentucky Core Content Test

Table of Contents

Summary

CTB and HumRRO independently calculated the scaled/linked raw score to scale score tables for the 1999 Kentucky Core Content Test. From those tables, both identified cutpoints that could be used for assigning student performance classifications and later converted to school accountability indexes. Differences between the calculations of CTB and those of HumRRO were small and typically did not affect student classification decisions. In the one instance where a cutpoint was different, the difference was only a single scale score point. Given that our scaling and linking results are either identical with CTB or within explainable tolerance of CTB, we can be assured that CTB did not commit processing errors.

Third-Party Checking of 1999 Scaling and Linking for the Kentucky Core Content Test

Introduction

In order to make the transition from the KIRIS test to the Kentucky Core Content Test with the minimum amount of disruption, a system of linking the old test with the new was necessarily devised. This link allows Kentucky to maintain consistency in its student performance levels and to apply the student Kentucky Core Content Test scores to a newly revised accountability calculation. The main difficulty in linking the two tests is that KIRIS only applied student scores on the open-response section of the test toward a school's accountability index and toward individual student performance levels. The Kentucky Core Content Test uses both open-response and multiple-choice format questions to make those determinations. Students still receive ratings in terms of the Novice, Apprentice, Proficient, and Distinguished levels of performance, but multiple-choice questions are now included in those determinations. A two-step process was used to make the link from the Kentucky Core Content Test back to the KIRIS scale on which student performance standards had been set in 1993. The first step involved analysis of 1998 data in which multiple-choice and open-response items were combined on a single scale and that combined scale equated to the open-response-only scale. HumRRO, in an earlier report (Hoffman, Thacker, & MacBride, 1999), performed a third-party evaluation of those procedures. The second step, for linking the Kentucky Core Content Test back to the KIRIS scale, was to link the 1999 test data to the newly created combined scale. This report represents a third-party evaluation of the procedures used in the second step.

Scaling and Linking Procedures

Scaling and linking procedures varied somewhat depending on the subject and grade. In all cases, 1999 item data from all forms were scaled using CTB's PARDUX program. Item parameters were then divided by form and entered into CTB's FLUX program to create raw score to scale score conversion tables. For Mathematics, Science, and Social Studies in all grades, and for Reading in Grades 4 and 7, the scaling process included adjusting item parameters by PARDUX application of the Stocking-Lord procedure to items linking Kentucky Core Content Test 1999 to KIRIS 1998. Anchor parameters for linking items were from the combined scaling of multiple-choice and open-response items conducted on the 1998 data. Cutpoints established in 1993 could then be applied to the 1999 scaling results. For Arts & Humanities, Practical Living/Vocational Studies, and Grade 10 Reading, cutpoints were established by equipercentile equating to 1998 performance level distributions.

Scope of Third-Party Checking

HumRRO conducted parallel analyses to accomplish scaling and linking for the 1999 data. Because of the severe time limits, HumRRO's analyses were constrained in three ways. First, CTB selected the calibration sample based on criteria set by KDE. HumRRO did not independently select a sample, but rather used the CTB selection. We did, however, check for missing data in the sample, a sign that students in the calibration sample may not have been completely processed by DRC. Second,

CTB conducted item-total raw score correlations on multiple-choice data and identified items to be excluded from scaling because of sufficiently poor item biserial correlations. HumRRO did not independently recalculate these correlations. Third, based on Stocking-Lord results, CTB iteratively identified abberant linking items that should be excluded from linking. HumRRO did not independently make decisions about which items to select, but did review the items selected by CTB while conducting Stocking-Lord analyses.

Processing Steps

HumRRO took the following steps for Mathematics, Science, and Social Studies in all grades, and for Reading in Grades 4 and 7:

1. Create anchor files (PARDUX *.anc) for specified multiple-choice test items that appear on both tests. These anchor items are used to link the 1999 test to the 1998 scale which was previously adjusted to the 1993 scale. A special SAS program was written to create this file, matching the linking items that bridged 1998 and 1999.

2. Prepare control files (PARDUX *.ctl) which contain the constraints used for item parameter estimation, student proficiency estimation, maximum number of items, etc. The SAS program used to create anchor files included a routine to print out a control file.

3. Create working files (PARDUX *.rwo) from the 1999 Kentucky Core Content Test calibration sample. These files include both open-response and multiple-choice data. Two different SAS programs were used to create rwo files. Because item placement in the rwo file is tedious, one program was used to generate lines of code that move input data into the correct output location. The second program applied those lines of code to the input files.

4. Estimate parameters for Kentucky Core Content Test items using PARDUX.

5. Perform Stocking-Lord transformation using PARDUX. The results of this transformation include a slope and intercept constant for linking the new 1999 Kentucky Core Content Test with the 1998 KIRIS test.

6. Confirm that the equating constants from Step 5 match those derived by CTB.

7. Create parameter files (FLUX *.par) for each test form for use in preparation of raw score to scale score tables. This was a “cut and paste” word processing task using PARDUX output of item parameters from step 4.

8. Create files (FLUX *.hlk) containing the scale limits (325 and 800) and constants from the Stocking-Lord transformation. This was a simple word processing task.

9. Create raw score to scale score transformation tables for each form using FLUX.

10. Confirm that the raw score to scale score transformation tables from Step 9 match those derived by CTB.

11. Confirm that the cutpoints set by CTB were consistent with established cutpoints from the KDE (1997) Cycle 2 Technical Manual and Wise (1998) Grade Shift Report.

Arts and humanities, practical living/vocational studies, and 10th grade reading tests were handled differently. Arts and humanities and practical living/vocational studies tests contained only half as many items as the other tests. These tests were not scaled for the 1998 data and therefore no linking was done for 1999. The high school reading test was moved this year from the 11th to the 10th grade. The shift in grade level in addition to the shift in testing procedures would have made equating for this test suspect. In order to link these tests to the 1998 KIRIS results, performance categories were determined using an equipercentile methodology. The proportion of students in the Novice, Apprentice, Proficient, and Distinguished categories in 1999 is defined by the number in 1998. This process involved:

1. Prepare control files (PARDUX *.ctl) which contain the constraints used for item parameter estimation, student proficiency estimation, maximum number of items, etc. As above, a SAS routine accomplished this task.

2. Create working files (PARDUX *.rwo) from the 1999 Kentucky Core Content Test calibration sample. These files include both open-response and multiple-choice data merged using the SAS routines described above.

3. Estimate parameters for Kentucky Core Content Test items using PARDUX.

4. Save student theta scores from PARDUX.

5. Create parameter (FLUX *.par) files for each test form by word processing of PARDUX output.

6. Create files (FLUX *.hlk) containing the scale limits (325 and 800) and constants for the score transformation (50 and 500 for each of these tests) by word processing.

7. Create raw score to scale score transformation tables for each form using FLUX.

8. Confirm that the raw score to scale score transformation tables from Step 6 match those derived by CTB.

9. Determine equipercentile cutpoints from cumulative frequencies of student thetas saved in step 4 and KDE-provided performance level distributions (See Table 1 below.)

10. Confirm that the cutpoints between the major categories, i.e., Novice to Apprentice, Apprentice to Proficient, and Proficient to Distinguished, match those derived by CTB.

Table 1
NAPD Performance Level Distributions for 1998

	Arts & Humanities			Practical Living/Vocational Studies			Reading
	Grade 5	Grade 7	Grade 11	Grade 5	Grade 7	Grade 11	Grade10
Novice	67.72	53.15	52.15	38.28	63.16	54.15	15.68
Apprentice	29.24	40.42	44.02	55.61	29.82	39.36	56.03
Proficient	0.66	4.16	1.90	4.87	4.97	4.25	26.46
Distinguished	2.37	2.27	1.93	1.24	2.05	2.38	1.83

Results

After performing periodic checks with CTB as individual tests were scaled and equated, HumRRO and CTB reached appropriate levels of agreement on the equating constants for mathematics, social studies, science, and fourth- and seventh-grade reading. The agreement level was typically very close, as indicated by Table 2. The largest discrepancy occurred in fifth-grade social studies, which coincides with the idiosyncrasy of the PARDUX program found during pre-equating (see Hoffman, Thacker, & McBride, 1999). Because of a communication error, HumRRO initially began the process with different transformation constants than did CTB, but ran the analysis both ways in order to demonstrate the magnitude of the difference in scale scores caused by the scaling discrepancy from the 1998 data.

The other significant difference between the results obtained by HumRRO and those from CTB is in 11th grade science. This difference is the result of using slightly different student raw score files. CTB used an earlier version of the student files that contained some missing values. HumRRO downloaded the file after the missing values had been added and, for comparison purposes, chose to analyze the more complete file rather than retrieve the older file with missing data. The differences caused by this discrepancy are very small and did not affect student classification into NAPD categories.

No discrepancies exist for arts and humanities, practical living/vocational skills, and Grade 10 reading. No equating constants were computed for these subjects. There was some discussion due to a small (about 1% of the students in the calibration sample) number of missing open-response data points in the sample for fifth- and eighth-grade arts and humanities. Students are scored zero if they leave an open-response item blank. Otherwise, they receive a score. DRC, the company that scores the tests, records a "B" for blank responses. Both HumRRO and CTB convert all "B's" to "0's" for equating. The scores in question were not "B" however, but actually missing. After computing different raw score to scale score tables due to handling these data differently, it was decided that missing data points should not be converted to "0's" and should instead be treated as "missing." HumRRO's and CTB's results match precisely using this rule. *(Note: In subsequent file updates, the missing data points were corrected, but initial scaling and linking results using the missing data were retained.)*

Early in the analysis there were discrepancies in the arts and humanities and practical living/vocational studies data due to some confusion about the convergence criteria used in PARDUX. For these subjects a convergence criterion of 0.01 was specified. Other subjects used a criterion of 0.005. A few of the early arts and humanities and practical living/vocational studies analyses were completed using the more stringent convergence criterion. The raw score to scale score tables computed using this criterion were slightly different than those computed using 0.01. It was determined that the differences were slight and would mean little in terms of altering students' and schools' scores. However, in order to maintain consistency, CTB reran those analyses and their results matched HumRRO's exactly.

Table 2 summarizes the results of this study. It identifies the grade and subject for each test in the first two columns. The next six columns contain the M1 and M2 (slope and intercept) constants obtained from the Stocking-Lord transformation. The first set of constants was computed using all anchor items. The second set was computed after turning off problem items (specified by CTB). In cases where no problem items were identified, the complete set of anchor items were used for the transformation. The remaining set of constants are those computed by CTB which were used for verification by HumRRO. From this data it can be surmised that the M1 and M2 constants computed by CTB closely match those computed by HumRRO.

Table 2 also contains columns that document comparison of raw score to scale score tables calculated by CTB and HumRRO. The first of these columns indicates the number of instances when the raw score to scale score computation is off by 1 scale score point for a given test. The second column indicates the number of instances when CTB's and HumRRO's raw score to scale score transformation differed by more than 1. As these columns indicate, CTB's and HumRRO's raw score to scale score tables never differed by more than 1 scale score point. Differences occurred in four grade/subjects. Grade 5 social studies was the test for which HumRRO and CTB used different anchor parameters. Data is also provided in the table (SS-CTB) for the computation using the same transformation constant as CTB. Correcting this discrepancy eliminates the difference in the raw score to scale score agreement. The Grade 11 science scaling was computed using slightly different sets of data. To maintain an analysis schedule that would meet the September 15 score release deadline, CTB used an earlier version of the calibration sample that had some missing data points. Despite a plan to use the same calibration sample, HumRRO downloaded a later, more complete, data set for Grade 11. The disparity produced by using the differing data sets is small and confirms that scaling does not need to be conducted with complete data sets. For the remaining two grade/subject combinations with raw score to scale score differences, that is, Grades 8 and 11 mathematics, the divergence appears to be due to slight differences in Stocking-Lord constants. In both of these cases, scaling was difficult and several items were deleted from linking (three in Grade 8 and five in Grade 11). Nevertheless, the divergence in raw score to scale score tables was did not affect the major NAPD classification points.

The last column of Table 2 refers to the agreement between CTB and HumRRO concerning the cutpoints for assigning the NAPD performance categories. Discounting the initial fifth-grade social studies computation, HumRRO and CTB agree exactly on the placement of the cutpoints for NAPD

categories in all but one instance. In 11th grade arts and humanities the CTB cutpoint for distinguishing between Novice and Apprentice on form 2B differs by one from HumRRO.

Table 3 contains a record of the checked cutpoints for each form of all tested subjects. Each column contains either a "Y" (matches exactly) or an "N" (does not match exactly) for each major cutpoint. Three major cutpoints were checked for each form, the cutpoint distinguishing between 1) Novice and Apprentice, 2) Apprentice and Proficient, and 3) Proficient and Distinguished. The only non-matching cutpoint was in form 2B of the 11th grade arts and humanities test. This cutpoint was different by 1 scale score point. Note that the matches occur only when using a rounding rule in which the scale being examined includes the interval up to .5 above that score (tabled scale scores are integers). When the cutpoint point is within some portion of the upper end of that scale score interval, then that scale score is counted as in the higher performance category. Table 3 identifies the Grade 11 arts and humanities cutpoint between Novice and Apprentice as the lone CTB cutpoint that differs from HumRRO.

Table 2
Comparison of HumRRO and CTB Scaling and Linking Results.

Grade	Subject	HumRRO Results				CTB Results		RS-SS Agreement		NAPD Agreement[1]
		All Anchor Items		CTB Set of Anchor Items						
		M1	M2	M1	M2	M1	M2	Off by 1[2]	Off by More than l	
04	RD	33.36016	545.53992	na	na	33.36007	545.53986	0	0	YES
04	SC	27.66099	540.12830	27.75393	539.76990	27.75424	539.76996	0	0	YES
05	MA	35.19292	553.62976	35.33308	553.01172	35.33280	553.01154	0	0	YES
05	SS-Hum[3]	32.49487	536.43005	31.61785	537.43182	31.60847	537.52130	41	0	[4]
05	SS-CTB	32.48417	536.51727	31.60847	536.51727	31.60847	536.51727	0	0	YES
05	AH	na	na	na	na	na	na	0	0	YES
05	PL	na	na	na	na	na	na	0	0	YES
07	RD	31.35150	511.29410	31.33557	511.37003	31.33554	511.37003	0	0	YES
07	SC	26.48189	499.24817	26.39925	499.30295	26.39926	499.30298	0	0	YES
08	MA	33.68691	528.43848	33.91278	527.61108	33.91141	527.60144	7	0	YES[5]
08	SS	37.83554	506.90787	38.38489	506.43484	38.38493	506.43460	0	0	YES
08	AH	na	na	na	na	na	na	0	0	YES
08	PL	na	na	na	na	na	na	0	0	YES
10	RD	na	na	na	na	na	na	0	0	YES
10	PL	na	na	na	na	na	na	0	0	YES
11	MA	39.95746	528.40735	39.84676	529.84772	39.84770	529.84668	1	0	YES[6]
11	SC[7]	29.60927	540.13318	31.10640	539.91016	31.11016	539.91016	38	0	YES[8]
11	SS	43.93030	543.00189	44.41350	543.55334	44.41317	543.55383	0	0	YES
11	AH	na	na	na	na	na	na	0	0	YES-1[9]

[1] Only the four NAPD categories were checked and not the high, middle, low subdivisions for N or A.
[2] Total possible agreements: 300 for all AH and PL; 511 for 05 MA, 05 SS, and 07 SC; 438 for the remaining.
[3] SS-Hum was run with results from the 50-cycle procedure for calculating anchor parameters. SS-CTB was run with results from the 25-cycle plus procedure for creating anchor parameters.
[4] For the 41 divergent raw score to scale score results, only one results in a divergent NAPD. In Form 6, raw score 27 would be an N with our results and an A with CTB's results.
[5] The seven divergent results do not affect basic four NAPD classifications.
[6] The divergent result does not affect basic four NAPD classifications.
[7] HumRRO used a later file which had more complete data for forms 1 and 2. Divergence represents a sensitivity test.
[8] The divergent results do not affect basic four NAPD classifications.
[9] CTB N/A cutpoint for Form 2B differs by 1 from HumRRO.

Table 3
Check on major cutpoints between Novice and Apprentice, Apprentice and Proficient, and Proficient and Distinguished[10]

		Form											
		1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B
Grade	Subject	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd	na/ap/pd
04	RD	YYY	YYY	YYY	YYY	YYY		YYY		YYY		YYY	
04	SC	YYY		YYY		YYY		YYY		YYY		YYY	
05	MA	YYY		YYY		YYY		YYY		YYY		YYY	YYY
05	SS	YYY		YYY	YYY	YYY		YYY		YYY		YYY	
05	AH	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY
05	PL	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY
07	RD	YYY		YYY		YYY		YYY		YYY		YYY	
07	SC	YYY		YYY	YYY	YYY		YYY		YYY		YYY	
08	MA	YYY		YYY		YYY		YYY		YYY		YYY	
08	SS	YYY		YYY		YYY		YYY		YYY		YYY	
08	AH	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY
08	PL	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY
10	RD	YYY		YYY		YYY		YYY		YYY		YYY	
10	PL	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY
11	MA	YYY		YYY		YYY		YYY		YYY		YYY	
11	SC	YYY		YYY		YYY		YYY		YYY		YYY	
11	SS	YYY		YYY		YYY		YYY		YYY		YYY	
11	AH	YYY	YYY	YYY	NYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY	YYY

[10] Agreement results assume a rounding rule – e.g., Scale score 509 includes the interval up to 509.5. For a cutpoint of 509.19, a scale score 509 is assigned to be in the higher category.

Documentation

To document the steps involved in scaling and linking the 1999 Kentucky Core Content Test we saved all electronic files used in data preparation, including SAS programs, SAS logs, and SAS output lists and all files produced during PARDUX scaling and FLUX transformations. These files have been submitted to KDE. To illustrate these files, this report includes a number of appendixes containing information for Grade 7 Science. File names indicated for Grade 7 Science can be used to decode file names of the remaining grade/subject combinations by substituting grade (e.g. 07) and subject (e.g., SC) indicators. The following are included:

Appendix

A. PARDUX Control File (Sc0799.CTL). This file contains the number of items, the maximum number of stages for PARDUX, the convergence criterion, parameter estimation limits, maximum and minimum values for proficiency estimates (theta), and other information. This file also contains information allowing the program to distinguish between open-response and multiple-choice items, the items to be calibrated, and the number of score levels for open-response data.

B. PARDUX Data File (Sc0799.RWO). This file contains the student score data. It is coded such that a 1 indicates a correct answer for a multiple-choice question and actual score levels (0-4) are recorded for student responses to open-response questions. To facilitate communication, HumRRO adhered to CTB's item order in constructing these data files.

C. PARDUX Anchor File (Sc0799.ANC). This file contains 1998 common-scaling item parameters for the 1999 Kentucky Core Content Test, Form 3 items. Aside from position, these items were unchanged from 1998 to 1999. Only multiple-choice items are used in *.ANC files.

D. SAS Programs for Creating Anchor Files, PARDUX Control Files and *.rwo (Working Data) Files. The run logs for four different programs are included. The first program (see Sc0798anc.log) assigns *.rwo locations for each item, matches *.rwo locations between 1998 and 1999 for linking items, creates anchor files from 1998 parameter files, and creates control files. The second (see SC0799CD.log) uses the rwo position information and writes lines of code that were then inserted into the third program. Note that Grade 7 Science was one of the grade/subject tests that required special treatment due to a difference on a operational item position in the A and B subforms. The third program (see Sc07rwo.log) merges multiple choice raw data and open response raw data to create an *.rwo file with items aligned according to CTB specifications. A fourth program (also see Sc07rwo.log) provides a partial comparison between the *.rwo file produces by HumRRO and the one produced by CTB.

E. PARDUX Parameter Estimation Summary (SC0799_SUM.TXT). This file provides a summary of the parameter estimation procedure run in PARDUX. It includes the limit data from the control file and also contains the number of stages PARDUX runs in order to reach convergence. It also contains the item numbers of items that could not be estimated and documents any items whose

estimation reaches the maximum alpha parameter. This file identifies any problem items that might require additional manipulation before continuing the process.

F. PARDUX Parameter Estimation Details (SC0799_DET.TXT). This file is a thorough iteration of the item data during each stage of parameter estimation. This file is typically quite large and is not included in its entirety in the appendices. Only the first few pages of the file are printed. These pages include the first stage of parameter estimation and the beginning of the second.

G. PARDUX Parameter File (Sc0798ch.PAR). This file contains parameter estimates for all items designated by the *.CTL file. It is used for later data manipulation.

H. PARDUX TST File (Sc0799.tst). This file can be used to calculate form reliabilities. It was created and saved for each grade and subject tested. No form reliabilities were calculated for this project.

I. PARDUX VEC File (Sc0799.vec). This file contains all student data and includes an estimation of proficiency for each student's score data. For arts and humanities, practical living/vocational studies, and 10^{th} grade reading this file was used to calculate cumulative frequencies. The cumulative frequencies were then used to calculate cumulative percents. That data was used to assign cut points to the data set such that an equal percentage of students were placed in each performance category as in 1998.

J. PARDUX Item Summaries Files, Status (SC0799_STAT.TXT). This file lists all items for a given test and their status after parameter estimation. Items are coded as either estimate OK, OK—default C, not estimated, or other codes. This file provides a different type of record for the parameter estimation.

K. PARDUX Item Summaries Files, Distribution (SC0799_DIST.TXT). This file contains the distribution of students who scored at each level on the open-response items. This file is useful for examining the way that scoring rubrics for these items operate and for ensuring that all open-response items have the correct number of functioning score levels.

L. PARDUX Item Summaries Files, Parameters (SC0799_PAR.TXT). This file contains the item parameters in more readily edited format than the *.PAR file. This file can easily be read into word processors and spreadsheet programs.

M. PARDUX Item Summaries Files, Standard Errors (SC0799_SE.TXT). This file contains the standard errors for each item including the errors for the various score levels on the open response items.

N. PARDUX Item Summaries Files, FitQ1 (SC0799_Q1.TXT). This contains fit statistics for all items.

O. PARDUX Log File (SC0799_LOG.TXT). As each manipulation of data is completed, PARDUX maintains a log of the procedures and filenames. This log is saved in text format.

P. Stocking-Lord Plots (SC0799_SLPLOTS.doc). The Stocking-Lord transformation of the data, which provides the M1 and M2 values (slope and intercept) that allow for the later creation of scoring tables outputs three graphs (one each for the a, b, and c parameters) for each transformation. In this file the three graphs that result from the transformation using all anchor items are included, as well as another set of similar graphs if any anchor items are deleted.

Q. Stocking-Lord Log (SC0799_SLLOG.TXT). This file contains a record of the Stocking-Lord procedures. It includes M1 and M2 values, both before and after any anchor items are omitted from the estimate.

R. FLUX control file (sc071999.HLK). This file specifies the range of the scale scores as well as the M1 and M2 transformation constants to be used from the Stocking-Lord transformation. An M1 of 50 and an M2 of 500 were used for those subjects/grades that were not equated in this manner.

S. FLUX Parameter Files by Form (SC0799FORM1.PAR, etc., one for each Form). Each of the parameter files computed using PARDUX was divided to represent items from each test form. Typically, 30 items were scored from each form. Arts and Humanities and Practical Living/Vocational Studies forms contained 10 items to be scored.

T. Raw Score to Scale Score Tables (SC0799_FLUX_RS.doc & SC0799FORM1_FLUX.TXT). A raw score to scale score table was produced for each form. These tables were saved in text format using FLUX for each individual form and also saved as a single large Microsoft Word document for each tested subject/grade. Appendix T represents the Word document; however, the .TXT documents are identical by form.

Conclusion

CTB and HumRRO independently calculated the scaled/linked raw score to scale score tables for the 1999 Kentucky Core Content Test. From those tables, both identified cutpoints that could be used for assigning student performance classifications and later converted to school accountability indexes. Differences between the calculations of CTB and those of HumRRO were small and typically did not affect student classification decisions. In the one instance where a cutpoint was different, the difference was only a single scale score point. Given that our scaling and linking results are either identical with CTB or within explainable tolerance of CTB, we can be assured that CTB did not commit processing errors.